



Cloning and sequencing of the Bet v 1-homologous allergen Fra a 1 in strawberry (*Fragaria ananassa*) shows the presence of an intron and little variability in amino acid sequence

Anna Musidlowska-Persson¹, Rikard Alm, Cecilia Emanuelsson*

Department of Biochemistry, Center for Chemistry and Chemical Engineering, Lund University, PO Box 124, S-221 00 Lund, Sweden

Received 13 February 2006; received in revised form 10 June 2006; accepted 12 June 2006

Abstract

The Fra a 1 allergen in strawberry (*Fragaria ananassa*) is homologous to the major birch pollen allergen Bet v 1, which has numerous isoforms differing in terms of amino acid sequence and immunological impact. To map the extent of sequence differences in the Fra a 1 allergen, PCR cloning and sequencing was applied. Several genomic sequences of Fra a 1, with a length of either 584, 591 or 594 nucleotides, were obtained from three different strawberry varieties. All contained one intron, with the length of either 101 or 110 nucleotides. By sequencing 30 different clones, eight different DNA sequences were obtained, giving in total five potential Fra a 1 protein isoforms, with high sequence similarity (>97% sequence identity) and only seven positions of amino acid variability, which were largely confirmed by mass spectrometry of expressed proteins. We conclude that the sequence variability in the strawberry allergen Fra a 1 is small, within and between strawberry varieties, and that multiple spots, previously detected in 2DE, are presumably due to differences in post-translational modification rather than differences in amino acid sequence. The most abundant Fra a 1 isoform sequence, recombinantly expressed in *Escherichia coli* after removal of the intron, was recognized by IgE from strawberry allergic patients. It cross-reacted with antibodies to Bet v 1 and the homologous apple allergen Mal d 1 (61 and 78% sequence identity, respectively), and will be used in further analyses of variation in Fra a 1-expression.

© 2006 Elsevier Ltd. All rights reserved.

Keywords: Allergen; Allergy; Birch pollen; Cross-reactivity; Mass spectrometry; Isoform

1. Introduction

Food allergies increase, and are usually triggered by the protein components in food (Breiteneder and Ebner, 2000; Mills et al., 2003). A small number of protein homologues are responsible for the majority of the allergies, with birch pollen associated food allergy dominating in northern Europe (Aalberse et al., 2001; Vieths et al., 2002). Specific IgE antibodies, raised against

inhaled birch pollen proteins, subsequently may cross-react with homologous proteins after ingestion of fruit, berries and vegetables.

As a challenge from plant breeder's point of view, it is expected that there is a biological variation in allergen content among fruit, berries and vegetables such that the allergenic impact may be different depending on variety (Hsieh et al., 1995; Vieths et al., 1994). Differences in allergenic impact between varieties may be due to differences in the total amount of a particular allergen protein, or to differences in allergen isoforms. For the major birch pollen allergen, Bet v 1, there are 10 different isoforms with amino acid sequence variability (Bet v 1A-L, >80% sequence identity) which are different in IgE binding and effect on T-cells reactivity (Ferreira et al., 1996). In apple, there are several isoforms of the Bet v 1-homologous major allergen Mal d 1, but so far it seems to be the total amount of Mal d 1 rather than isoform variation that determines the allergenic impact (Marzban et al., 2005; Son et al., 1999). For cherry, it was recently described that the Bet v 1-homologous major allergen

Abbreviations: 2DE, 2-dimensional polyacrylamide gel electrophoresis; pI, isoelectric point; rMal d 1, recombinantly expressed Mal d 1; rBet v 1, recombinantly expressed Bet v 1; His-rFra a 1, His-tagged recombinantly expressed Fra a 1; nFra a 1, native Fra a 1; Mal d 1, major allergen in apple (*Malus domestica*); Bet v 1, major allergen in birch (*Betula verrucosa*) pollen; Fra a 1, major allergen in strawberry (*Fragaria ananassa*); EST, expressed sequence tag, partial DNA sequences obtained from cDNA; nt, nucleotide; MS, mass spectrometry; MS/MS, tandem MS

* Corresponding author. Tel.: +46 46 222 48 72; fax: +46 46 222 45 34.

E-mail address: Cecilia.Emanuelsson@biokem.lu.se (C. Emanuelsson).

¹ Present address: Syngenta Seeds AB, Box 302, 261 23 Landskrona, Sweden.

Pru av 1 appeared in five different spots in two-dimensional gel electrophoresis (2DE), with all five spots reacting with IgE from cherry-allergic patients (Reuter et al., 2005). However, whereas four different isoforms of Pru av 1 were obtained by screening a cDNA library, all these four different isoforms obtained from cDNA were detected in all five spots in 2DE, thus no differences were found that could explain why Pru av 1 appears in the five different spots.

In strawberry (*Fragaria ananassa*) we have identified a Bet v 1-homologous allergen, Fra a 1, utilizing IgE from patient samples for detection and protein mass spectrometry for characterization of the allergen (Karlsson et al., 2004). Furthermore, proteins extracted from strawberry were resolved as 1000 different spots in 2DE, with the Fra a 1 allergen detected in four spots in the 20 kDa region between pI 5.7 and 6.0 (Hjernø et al., 2006). However, when analyzing the spots by mass spectrometry, we could detect very little variation in the amino acid sequence of Fra a 1. We found a A142S-substitution in one peptide, and in a second peptide a L114V substitution, a G112 deletion and a V24L substitution, but as in the above-mentioned cherry study (Reuter et al., 2005), these variants were detected in all spots, and none of these amino acid substitutions could explain a shift in pI. However, the total sequence coverage for Fra a 1 was only 75%, which leaves open a possibility that there could be differences in amino acid sequence that could explain the shift in pI between the spots in the parts of the amino acid sequence not covered by the mass spectrometric analysis.

To make an estimation of the amino acid sequence variability by an independent method, we therefore designed primers, based on the partial sequence information obtained by mass spectrometry and matching to EST-sequences (Hjernø et al., 2006), and here describe PCR-based cloning and sequencing of the strawberry allergen Fra a 1. Several genomic sequences with a length of 584–594 nt were obtained, from three different strawberry varieties, all containing one intron with the length of 101 or 110 nt. In total, five potential Fra a 1 protein isoforms were suggested from sequencing 30 different clones. Only seven positions of variability were indicated in the deduced amino acid sequence. For three of these both the alternative amino acids are detected by mass spectrometry (Hjernø et al., 2006).

We conclude from cloning and sequencing, corroborated by mass spectrometric analysis of expressed protein, that the amino acid sequence variability in the strawberry allergen Fra a 1 is small, and that the appearance of multiple spots in 2DE is due not to differences in the amino acid sequence but rather to post-translational modification, the identity of which remains to be determined. Another important conclusion is that the comparison of three different varieties of strawberry suggests that different varieties have similar, rather than different, setups of allergen isoforms. We have also expressed the major Fra a 1 isoform as recombinant protein in order to develop specific antibodies and affinity enrichment of native Fra a 1 protein directly from strawberry, to assess its variation, in different strawberry varieties and growth conditions, in terms of quantity and post-translational modification.

2. Materials and methods

2.1. Cloning and DNA sequencing

Chromosomal DNA was isolated from leaves of three different varieties of strawberry (*Fragaria ananassa* var. Korona, *Fragaria ananassa* var. Honeoye, and *Fragaria ananassa* var. X, an unknown, locally available variety) according to the GenElute Plant Genomic DNA Kit (Sigma, Steinheim, Germany) and used as template in PCR. The PCR program used was 95 °C for 2 min (1 cycle); then 95 °C for 15 s, 53 °C for 30 s, 68 °C for 1 min (35 cycles) and finally 68 °C for 10 min (1 cycle). The reaction mix (total volume of 30 µl) contained 1 U of AccuPrime Pfx DNA polymerase (Invitrogen, Carlsbad, USA) and 3 µl reaction buffer, 0.3 µM of each primer (MWG-Biotech, Ebersberg, Germany) and 2 µl of isolated genomic DNA solution. The primers used were attB1_1f (GGG GAC AAG TTT GTA CAA AAA AGC AGG CTC TTA TTT CTC AAA TGT CGT GAT CAT CAT G), attB2_1r (GGG GAC CAC TTT GTA CAA GAA AGC TGG GTG TAC TCA AAC ATC ATA TAG AGA A(A/C)G ATC AT). The PCR-product was gel purified and cloned into a pDONR201 vector using BP reaction of GATEWAY (Invitrogen, Carlsbad, USA) cloning method in accordance with the manufacturer's instructions. DNA-sequencing was carried out on plasmid DNA at MWG-Biotech (Ebersberg, Germany).

2.2. Intron removal

Splicing by overlap extension (Horton et al., 1993) was used to remove the intron from the gDNA in order to obtain the coding region for expression of Fra a 1 protein in *Escherichia coli*. The reaction was carried out in one step using AccuPrime Pfx DNA polymerase (Invitrogen, Carlsbad, USA) with pDONR201, containing the genomic DNA sequence of Fra a 1, as template and two primers: Intron_f GAA GAT CAC CTT TGG TGA AGG CAG CCA CTA CGG GTA CGT C and Intron_r GAC GTA CCC GTA GTG GCT GCC TTC ACC AAA GGT GAT CTT C. The PCR program used was 95 °C for 2 min (1 cycle); then 95 °C for 15 s, 55 °C for 30 s, 68 °C for 4 min (35 cycles) and finally 68 °C for 10 min (1 cycle). The product of the reaction was expected to be a circular pDONR201 containing the coding sequence of Fra a 1. After the PCR-reaction the mixture was treated with 1 U of DpnI enzyme in order to digest the template plasmid and transformed into *E. coli* DH5α cells. The resulting colonies were checked for presence of the Fra a 1 coding sequence by colony PCR. To perform this a colony was picked, suspended in 100 µl of H₂O and boiled at 100 °C, 10 µl of this suspension was used as template for PCR using *Nde*I-Fraa1_f and *Bam*HI-Fraa1_r oligonucleotides as primers. The correct DNA sequence was confirmed by sequencing.

2.3. Expression of recombinant Fra a 1

The first step in the expression vector construction was the removal of the intron. The splicing by overlap extension method (Horton et al., 1993) was used to get the coding sequence of Fra a 1. The reaction was performed in one step using the

pDONR201 enclosing the genomic sequence of the allergen as a template. The intron was spliced away giving a circular product pDONR201-rFraa1. The coding sequence was subcloned into the pET-15b vector. Since the *E. coli* codon usage analysis with the allergen coding sequence showed presence of three rare proline CCC and seven glycine GGA codons, *E. coli* Rosetta DE3 harbouring a plasmid encoding tRNAs for those codons was chosen as an expression strain. In order to obtain the expression vector the gene was amplified with two primers introducing a *NdeI* site (*NdeI*-Fraa1.f GGA ATT CCA TAT GGG TGT TTA CAC TTA TGA A) and an *BamH* I site (*BamH*I-Fraa1.r CGC GGA TCC TTA GTT GTA TTC GCT GGG) at its 5'-end and 3'-end, respectively. *NdeI* and *BamH*I digestion of the resulting amplification product and ligation into the likewise-cut plasmid pET-15b (Novagen-EMD Biosciences, Darmstadt, Germany) obtained a pET-15b-Fraa1 vector, which encodes for a (His)₆-tag and a thrombin cleavage site, which was transformed into *E. coli* DH5- α cells. Positive clones were selected by PCR screening and the Fra a 1 sequence was controlled by sequencing. After purification, the positive vectors were used for transformation of *E. coli* Rosetta DE3 strain (Novagen-EMD Biosciences, Darmstadt, Germany). The 100 ml LB broth supplemented with ampicillin (100 μ g/ml) and chloramphenicol (50 μ g/ml) was inoculated with a fresh overnight culture (the initial OD₆₀₀ of 0.07) and grown at 37 °C, protein expression was induced at OD₆₀₀ around 0.6–0.7 by addition of isopropyl- β -thio-galactopyranoside (IPTG) to a final concentration of 1 mM. After 3 h of culture, the cells were harvested by centrifugation (3220 $\times$ g, 4 °C, 10 min). The cell pellet was washed once and resuspended in 10 ml phosphate buffer (sodium phosphate buffer 100 mM, pH 7.5), incubated with lysozyme (100 μ g/ml) at 37 °C for 15 min and then sonicated to disrupt the cells. After centrifugation (3220 $\times$ g, 4 °C, 25 min) the supernatant was stored in –20 °C.

2.4. Electrophoresis and immunoblotting

E. coli cell extracts containing His-rFra a 1 (5 μ l) were analysed by SDS-PAGE on polyacrylamide gels (12%, BioRad, Hercules, USA). As reference to estimate the molecular mass of the bands, the Precision Plus Protein Standards (BioRad, Hercules, USA) were used. Gels were stained for protein detection with Coomassie brilliant blue. Blotting transfer was performed on to a PVDF membrane (Micron Separations Inc., Boston, USA) by a semidry blotter according to (Bjerrum and Schafer Nielsen, 1986; Harndahl et al., 1998). As primary antibody rabbit anti-Mal d 1 or rabbit anti-Bet v 1 antibody at 1:1000 dilution was used. The secondary antibody was alkaline phosphatase-conjugated goat anti-rabbit IgG antibody (BioRad) at 1:3000 dilution. The protein bands were visualized after reaction with BCIP and NBT (Pierce, Rockford, USA) based on phosphatase activity in accordance with the manufacturer's instructions. Immunodetection based on IgE reaction was performed similarly using patient serum at a 1:10 dilution as primary and alkaline phosphatase-labelled murine monoclonal anti-human IgE antibody (DPC, Los Angeles, USA) at a 1:5000 dilution as secondary antibody.

2.5. Mass spectrometry

The band representing the His-tagged rFra a 1 was excised from SDS-PAGE gel and subjected to in-gel-digestion and peptide extraction (Hjermø et al., 2006). The dried peptides were redissolved in 0.1% TFA and the solution obtained was treated with to convert lysines into homoarginines (no arginines are present in the strawberry allergen). Ten microliter of the extracted peptides solution was diluted with 10 μ l H₂O, 10 μ l of ammonium hydroxide (30%, v/v) and 5 μ l of guanidination reagent (50 mg of *O*-methylisourea hemisulfate dissolved in 51 μ l H₂O) were subsequently added, each step followed by vortexing. The sample was incubated at 60 °C for 12 min and the reaction mixture was directly loaded on the C18 STAGE microcolumn (Rappsilber et al., 2003) – C18 Empore Disks (3 M, Minneapolis, USA) packed in a GELoader tip (Eppendorf Scientific, Westbury, USA) – for desalting and concentration, as described previously (Gobom et al., 1999). Peptides were eluted from the microcolumn with 0.75 μ l matrix solution (5 mg/ml α -hydroxy cinnamic acid in 70% (v/v) CH₃CN and 0.1% TFA) directly onto the MALDI target. Manual data acquisition was performed using an Applied Biosystems 4700 Proteomics Analyzer with TOF/TOFTM optics (Applied Biosystems, Darmstadt, Germany), in positive reflector mode. Laser intensity and number of laser shots were varied depending on the peptide analyzed. In the manual setup, an MS-spectrum was first recorded and analyzed manually by a first round of peptide mass search. Peaks were selected for MS/MS analysis based on the obtained results. MS/MS spectra were included in the analysis and the total scores reported.

2.6. Recombinant allergens and antisera

Recombinantly expressed major allergens from birch and apple, rBet v 1 and rMal d 1, respectively and polyclonal anti-Bet v 1 and anti-Mal d 1 antisera were obtained for non-commercial use from ALK-Abello, Hørsholm, Denmark.

3. Results

3.1. Cloning and sequencing of 30 Fra a 1 genomic DNA sequences from three different strawberry varieties

Preliminary sequence data for Fra a 1 were obtained by peptide sequencing with mass spectrometry and matching data to nucleotide and EST databases (Hjermø et al., 2006). At least two EST-sequences (contig_Fra and jb125) were found to match MS data representing the allergen, and another nucleotide sequence (GenBank accession no. AY679601) was discovered by a homology search. These nucleotide sequence data were used to design the PCR primers for cloning of Fra a 1 gDNA from strawberry. The primers were designed to be complementary to the up- and downstream regions of the Fra a 1 gene in the contig_Fra sequence (Fig. 1), and also included flanking regions for the Gateway cloning system.

These primers were used for amplification of Fra a 1 sequences from gDNA isolated from strawberry leaves, from

```
Contig_Fra      GCACGAGGGTCCCTTATTCACGAGGCTTCTTATTTCTCAAATGTCGTGATCATCATGGG 60
jb125          -TCTTAGCTTCCATAT--ATATAGTTGTCTTATTTCTC-----TGATCATCANGGG 48
                **  ***  ***  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *

Contig_Fra      AAGGACCACCCAGCGAATACAACCTAAGTTATGATCGTTCTCTATATGATGTTTGAGTAC 571
AY679601        AAGGACCACCCAGCGAATACAACCTA-GTTATGATCTTCTCTATA----- 242
                ***** *****
```

Fig. 1. Primer design based on sequence information obtained on the strawberry allergen Fra a 1 by mass spectrometric peptide sequencing and fitting data to EST-sequences. The sequences used in primer design, shown on grey background in the Contig_Fra sequence, were complementary to the up- and downstream regions of the Fra a 1 gene. The nucleotides in the Contig_Fra sequence, built from several EST-sequences, were confirmed in most positions by identity with nucleotides from two sequences from other sources, the jb125 sequence (N-terminal) or the GenBank accession no. AY679601 sequence (C-terminal), as described in Hjerno et al. (2006). For the position G547T degenerate primers were designed.

the strawberry variety Korona, previously investigated by mass spectrometry (Hjerno et al., 2006), and two other varieties. The PCR products were inserted into the donor vector pDONR201 by a BP recombination reaction of Gateway system. Thirty clones were sequenced giving 11 different DNA sequences. Only the DNA sequences which have been identified from at least two clones or single sequences giving the same deduced amino acid sequence (like e1 and e2, Fig. 2) were considered as correct ones. In this way eight DNA sequences with a length of 584, 591 or

594 nt were obtained (Fig. 2). The sequence identity among the sequences was high (>97%), and their distribution among the 30 colonies, obtained from three different strawberry varieties are summarized in Table 1. The deduced amino acid sequence represented by a1, a2, b1, b2 (=same amino acid sequence except V39I and V114I substitutions) was detected in 59% of the clones and in all three varieties. The sequence designated “a”, submitted as GenBank accession no. GI:74197561, is UniProtKB/TrEMBL entry Q3T923.

```
a1 ATGGGTGTTTACACTTATGAAAACGAGTTCACCTCCGATATCCAGCACCTAAGTTGTTCAAGGCCTTCGTTCTCGATGCCGACAACTCATCCCAAGATTGCACCTCAGGCAGTAAAG 120
a2 .....T..... 120
b2 .....A..... 120
d .....G..... 120
b1 .....A..... 120
c .....T..... 120
e1 .....T..... 120
e2 .....T..... 120
*****

a1 TGCGCCGAGATCTCGAAGGAGATGGCGGCCCGGAACCATCAAGAAGATCACCTTTGGTGAAGGTATACT-----CTCAATCCATTTTACGTATATTTCAATTTGGTAATTT 230
a2 .....C..... 230
b2 .....C..... 240
d .....C..... 240
b1 .....C..... 240
c .....C..... 240
e1 .....C..... 240
e2 .....C..... 240
*****

a1 TTGGACCGTAGTAAATTATATGCTCACTCCCTTTTTTACATTGGAACCTCGACAGGCAGCCACTACGGGTACGTCAGGCACAAAGATCCACTCTATTGACAAAGAGAACCACACCTACAGC 350
a2 .....C..... 350
b2 .....G..... 360
d .....G..... 360
b1 .....G..... 360
c .....G..... 360
e1 .....T..... 360
e2 .....C..... 360
*****

a1 TACAGTTTGATCGAGGGAGATGCCCTGTGTCAGACACATCGAGAAGATCGACTACGAGACCAAGTTGGTATCAGCTCCTCACGGAGGAACCGTCAATCAAGACCACAGCAAGTACCACACC 410
a2 .....A..... 410
b2 .....A..... 420
d .....A..... 420
b1 .....A..... 420
c .....G..... 420
e1 .....T..... 420
e2 .....G..... 420
*****

a1 AAGGGAGATGTCGAGATCAAGGAAGAGCACGTCAGGCCGCGCAAGAGAAGGCCTCGCACCTCTTCAAGCTCATTGAGGGATACCTCAAGGACCACCCAGCGAATACAACCTAA 584
a2 .....A..... 584
b2 .....A..... 594
d .....A..... 594
b1 .....A..... 594
c .....G..... 591
e1 .....G..... 594
e2 .....G..... 594
*****
```

Fig. 2. Multiple DNA sequence alignment of eight different sequences of the strawberry allergen Fra a 1. In total 30 clones were sequenced giving 11 different DNA sequences, but only the DNA sequences that gave more than one identical sequence, in terms of either nucleotide sequence or deduced amino acid sequence, were considered as correct ones, giving in total eight DNA sequences designated a1, a2, b1, b2, c, d, e1 and e2, where, e.g. a1 and a2 represent the same deduced amino acid sequence. All sequences contained an intron (grey shaded). The frequency of the different sequences is presented in Table 1.

Table 1

Frequency of the eight different DNA sequences observed by cloning and sequencing Fra a 1 from three strawberry varieties

Sequence (see Fig. 2)	GenBank accession number	<i>Fragaria ananassa</i> variety Korona	<i>Fragaria ananassa</i> , variety Honeoye	<i>Fragaria ananassa</i> variety unknown	Intron deletion	Gly112 deletion	Protein
a1	DQ385511, gi 88082484	5	–	1	+		Fra a 1-A
a2	AM084674, GI:74197561	1	–	3	+		Fra a 1-A
b1	AM236313, gi 90185677, AM236314, gi 90185679	5	1	–			Fra a 1-B
b2	AM236315, gi 90185681	–	–	2			Fra a 1-B
c	AM236317, gi 90185685, AM236318, gi 90185687	2	3	–		+	Fra a 1-C
d	AM236316, gi 90185683	–	–	2			Fra a 1-D
e1	AM236319, gi 90185689	1	–	–			Fra a 1-E
e2	AM236320, gi 90185691	1	–	–			Fra a 1-E

Cloning and sequencing of 30 colonies gave eight DNA sequences designated a1, a2, b1, b2, c, d, e1 and e2 as outlined in Fig. 2, and with the sequence distribution among the colonies (15 colonies from variety Korona, 12 colonies from the other two varieties) summarized below. In addition three sequences f, g, h were detected but only once, and therefore not submitted to database. The deduced amino acid sequence represented by a1, a2, b1, b2 (the same amino acid sequence except substitutions at V39I and V114I) was detected in 59% of the clones and in all three varieties. The sequence represented the deduced amino acid sequence Fra a 1-A is appearing as UniProtKB/TrEMBL entry Q3T923. Here described cloned and sequenced five isoform amino acid sequences (A–E) appear as UniProtKB/Swiss-Prot entries Q3T923, Q256S7, Q256S6, Q256S2, and Q256S4.

3.2. Presence of intron

There was one intron (101 or 111 nt) present in all sequences. The length of the intron was deduced according to the literature data considering the splicing sites for plant genes (Goodall and Filipowicz, 1991) and confirmed using the Fra a 1 amino acid sequences obtained by mass spectrometry (Hjernø et al., 2006). Of investigated 30 clones, 10 possessed a nucleotide-deletion in the intron. The sequence variant with this deletion of 10 nt in the intron was detected in two of the three varieties.

3.3. Comparison of data from DNA sequencing and protein mass spectrometry

In the exon regions, variability was detected in seven positions in the deduced amino acid sequence (position 3, 39, 78,

95, 112, 114, 142), indicating the existence of at least five isoforms of the Fra a 1 protein (Fig. 3), amino acid sequences (A–E) appearing as UniProtKB/Swiss-Prot entries Q3T923, Q256S7, Q256S6, Q256S2, Q256S4. In three of these seven positions, both the alternatively suggested amino acids have been detected by mass spectrometry (G/–112, V/I114 and S/A142), and in the other four positions V/G3, V/I39, E/V78 and D/E95, one of the amino acids were detected by mass spectrometry (V3, V39, V78, E95) (Hjernø et al., 2006). Only in one position, position 24, mass spectrometry detected two alternative amino acids, Val 24 and Leu24, whereas the cloned sequences only suggested Leu24.

The two different approaches, sequencing of Fra a 1 genomic DNA and mass spectrometry of the Fra a 1 protein, thus essentially confirm each other concerning that few positions were showing variability in the Fra a 1 amino acid sequence.

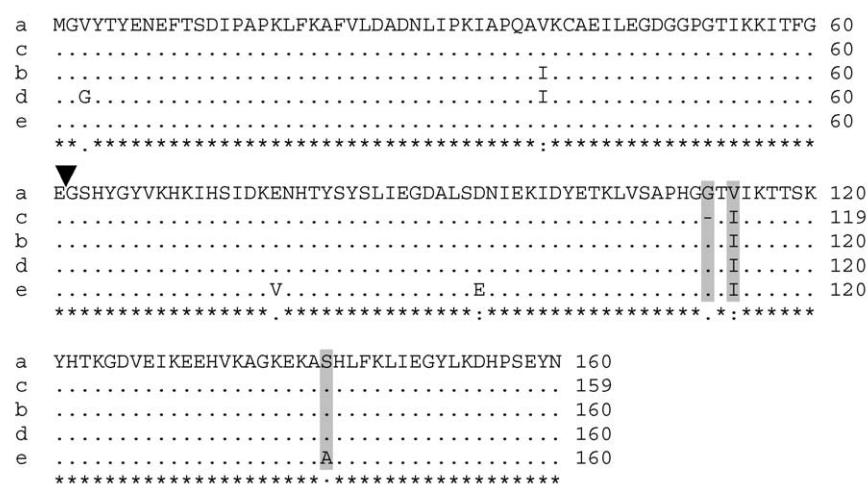


Fig. 3. Multiple amino acid sequence alignment of the five isoforms of the strawberry allergen Fra a 1. The eight DNA sequences designated a1, a2, b1, b2, c, d, e1 and e2 described in Fig. 2 gave five different variants of deduced amino acids designated a–e. The deduced amino acid sequence represented by a1, a2, b1, b2 (=same amino acid sequence except a V39L substitution) was detected in 59% of the clones, and in all three varieties as shown in Table 1. The sequence designated “a” has been submitted as GenBank accession nos. AM084674; GI:74197561, appearing as UniProtKB/TrEMBL entry Q3T923 (previous entry Q5ULZ4 is only a partial sequence). The here described cloned and sequenced five isoform amino acid sequences (A–E) appear as UniProtKB/Swiss-Prot entries Q3T923, Q256S7, Q256S6, Q256S2, Q256S4. Between these five isoforms, seven amino acid positions showed variability. In all these positions, at least one of the amino acids was detected by mass spectrometry, and in three of the positions (grey shaded), both the amino acid alternatives were detected. Arrowhead shows location of the intron.

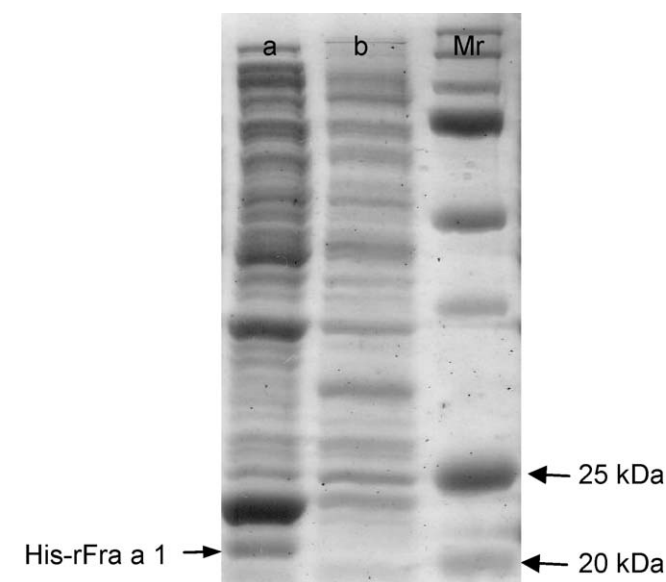


Fig. 4. Recombinant expression of the strawberry allergen Fra a 1. After removal of the intron, the most common isoform (sequence a in Fig. 3) was expressed in *E. coli*. SDS-PAGE (CBB-stained) shows proteins detected in 5 μ l of the bacterial extract from cells transformed (a) and not transformed (b) with the pET-15b-rFraa1 vector; Mr, molecular weight standard. The identity of the His-tagged recombinantly expressed Fra a 1 protein (band marked His-rFra a 1) was confirmed by MALDI mass spectrometry as described in Fig. 5.

3.4. Recombinant expression of Fra a 1

As seen in Fig. 2 and Table 1, the isoform that dominates is “a” (33% of the clones, b in 26%, c in 16%, d and e, each in 6% of the clones). The sequence designated “a” (GenBank accession no. AM084674; GI:74197561, UniProtKB/TrEMBL entry Q3T923) was used for construction of the expression vector in order to produce recombinant Fra a 1 protein after removal of the intron. The soluble fraction of the *E. coli* extract analyzed by SDS-PAGE showed two strong, inducible protein bands around 20–25 kDa, where the recombinant protein with calculated molecular weight, 19.947 kDa, was expected (Fig. 4). To identify the rFra a 1, samples were excised for MS analysis. The lower band of the two inducible bands contained the recom-

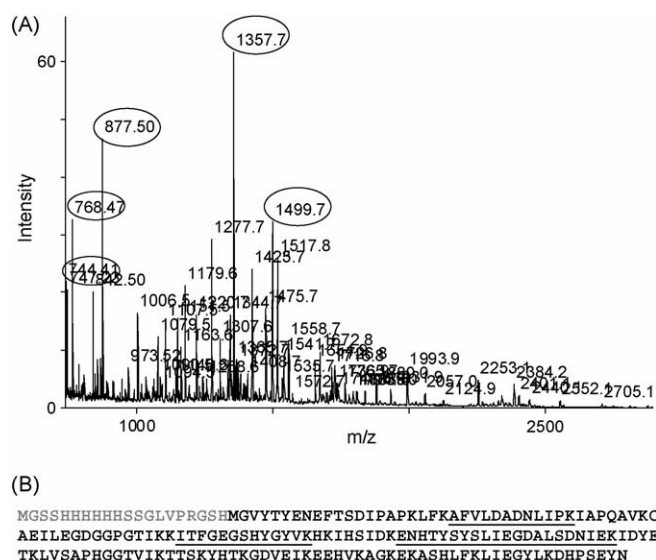


Fig. 5. MALDI-TOF mass spectrum of tryptic digest obtained from recombinantly expressed strawberry allergen Fra a 1. To confirm the identity of the recombinantly expressed rFra a 1 samples were excised from the SDS-PAGE in Fig. 4. (A) MALDI mass spectrum obtained after derivatization of the peptides with *O*-methylisourea for conversion of C-terminal lysines into homoarginines ($\Delta 42$ Da). Five abundant peptides derived from the rFra a 1 are marked by encircled peptide mass value. (B) Sequence of rFra a1 including the added His-tag and thrombin cleavage site (grey letters), with peptides detected by MALDI-MS underlined. The peptides aa 22–33 AFVLDADNLIPK, aa 57–69 ITFGEGSHYGYVK and aa 147–153 LIEGYLK were also confirmed by MS/MS (data not shown).

binant Fra a 1 allergen, as shown with 52% sequence coverage and 10 recognized peaks (Fig. 5), of which three were confirmed by MS/MS. The upper band of the two inducible bands proved to be the chloramphenicol acetyltransferase, which confers the antibiotic resistance (data not shown).

To explore the cross-reactivity of Fra a 1 with antisera against apple and birch pollen homologues, immunoblotting was performed, showing that the recombinantly expressed Fra a 1 was readily detected with both these antisera (Fig. 6A). The native protein, nFra a 1, in a protein extract from strawberry (expected mass 17.8 kDa) showed a characteristic doublet band (Karlsson

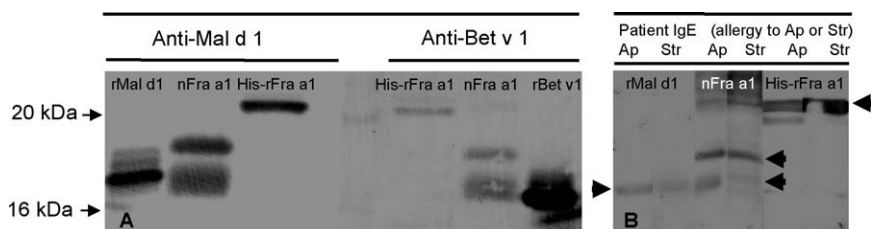


Fig. 6. Immunoblotting of recombinantly expressed strawberry allergen Fra a 1. (A) Cross-reactivity between recombinantly expressed strawberry allergen Fra a 1 and polyclonal antisera against the apple and birch pollen homologues Mal d 1 and Bet v 1: immunoblotting, utilizing as primary antibody either rabbit anti-Mal d 1 or anti-Bet v 1, of recombinantly expressed Fra a 1 (His-rFra a 1, 5 μ l/lane of the *E. coli* cell extract) compared to the native protein (nFra a 1, 50 μ g/lane of total protein extract from strawberry var. Korona) and recombinantly expressed and purified apple and birch pollen homologues (rMal d 1 and rBet v 1; each 0.3 μ g/lane). Molecular mass indicators to the left are estimated from the location of two of the bands in BioRad Precision Plus Standard proteins (20, 15 kDa) not visible in this image. (B) Immunodetection of native Fra a 1 and recombinantly expressed strawberry allergen Fra a 1 by human IgE in patient sera: immunoblotting, utilizing as primary antibody either a pool of sera from 15 patients allergic to apple (Ap) or serum from one patient allergic to strawberries (Str, patient number 2; Karlsson et al., 2004) and as secondary antibody goat anti-human IgE. The recombinantly expressed Fra a 1 (His-rFra a 1; 5 μ l/lane of the *E. coli* cell extract) is compared to the native protein (nFra a 1; 50 μ g/lane of total protein extract from strawberry var. Korona) and recombinantly expressed and purified apple pollen homologue (rMal d 1; 0.3 μ g/lane). Arrowheads high-light locations of the bands corresponding to rMal d 1, nFra a 1 (doublet) and His-rFra a 1.

et al., 2004). The explanation to the doublet appearance is yet not clear but it is described also for the Pyr c 1-homologue (major pear allergen, Karamloo et al., 2001) and is presumed to be some conformer phenomenon. The His-rFra a 1 band is located just above the 20 kDa marker, at a higher mass than the nFra a 1-doublet, which should be expected due to the mass increase of the attached tag. After thrombin cleavage to remove the tag, rFra a 1 runs as an 18 kDa protein, corresponding to the calculated mass of 18054.16 Da (data not shown). The reaction with anti-Mal d 1 is stronger than with anti-Bet v 1, probably due to the larger sequence similarity between Fra a 1 and Mal d 1 (77%) compared to Bet v 1 (55%).

To ensure that the recombinant Fra a 1 protein also reacted with human IgE, immunoblotting was performed with anti-human IgE as secondary antibody, utilizing as primary antibody either a serum from an individual with strawberry allergy (patient number 2 (Karlsson et al., 2004)), or a serum pool from 15 different individuals with apple-allergy (Fig. 6B). With both, a positive reaction occurred with rMal d 1, with nFra a 1 in strawberry protein extract and with His-rFra a 1 in *E. coli* extract, confirming that human IgE reacts with the recombinantly expressed Fra a 1-protein. It is not clear whether it is significant that the rMal d 1 is detected at the position of the lower band in the nFra a 1 doublet, nor why the pooled sera from patients with apple allergy recognized the lower nFra a 1 band better than the serum from the strawberry-allergic patient, which mainly recognized the upper nFra a 1 band.

4. Discussion

By PCR based cloning and sequencing, combined with analysis by protein mass spectrometry, we have shown that there appears to be relatively little sequence variability in the strawberry allergen Fra a 1. Sequence information from 30 genomic clones suggested only five different isoforms of Fra a 1, all sharing a high degree of sequence similarity (sequences b, c, d and e showed 98.8, 98.8, 98.1, and 97.5% sequence identity with sequence a). All sequences contained the intron position at codon 62 which is well-conserved among Bet v 1-homologous proteins (Hoffmann-Sommergruber et al., 1997).

Three different varieties of strawberries were used for cloning and sequencing. Small, differences in the frequency of distribution of the sequence variants a–e were observed among the colonies derived from the three strawberry varieties (Table 1). Yet, the >97% sequence identity among a–e and the consensus amino acid sequence (Q3T923 in UniProtKB/TrEMBL), if accounting for V–L substitutions at positions 39 and 114, detected in all three varieties gives the overall impression that there is relatively little variability in the amino acid sequence of the strawberry allergen Fra a 1, not only within but also between strawberry varieties. If different varieties of strawberry have similar, rather than different, setups of allergen isoforms as suggested by the here presented data, screening for hypoallergenic strawberry varieties could be limited to assessment of amount of Fra a 1-allergen rather than Fra a 1 isoform differences.

The sequence variability among the Fra a 1 sequences in strawberry appears surprisingly small compared to the recent

investigation of apple with genomic cloning and linkage mapping (Gao et al., 2005), where also several intronless sequences were recovered and a much lower degree of sequence identity was observed among the cloned sequences (60–70%). Yet there is a larger possible allelic variation in an octaploid species like strawberry with 16 alleles, compared to the diploid apple, and more varieties of strawberry (three) were investigated here, compared to only two apple varieties. Nevertheless, data obtained seem to indicate that there is less amino acid sequence variability in the Bet v 1-homologous allergen in strawberry, Fra a 1, compared to the Bet v 1-homologous allergen in apple, Mal d 1. However, maybe not all Mal d 1 sequences detected by cloning are expressed as proteins. It is known that intron-containing members show sequence similarity with Bet v 1 isoforms constitutively expressed in mature birch pollen, whereas intron-less sequences show sequence similarity with pathogen-inducible Bet v 1-like proteins (Swoboda et al., 1995a,b). Alternatively, strawberry Fra a 1 is indeed less heterogeneous than apple Mal d 1. Strawberry as a species, created by Duchesne around 1764, has existed only for a few 100 years, and cultivated strawberry maybe is not very outbred since it is often multiplied by cloning rather than crossing.

Could a larger sequence variability in Fra a 1 have been detected if more degenerate primers had been used? We tested one other entirely different primer pair (complementary to regions inside the Fra a 1 sequence, instead of up- and downstream regions) that also did not yield more than a single PCR-product and hence no sign of any intronless Fra a 1 sequence (data not shown), but except this we essentially used only one pair of primers for all cloning (with degenerate primers only to cover the position G547T in Fig. 1). However, both the approaches used here to investigate Fra a 1, DNA cloning and mass spectrometry of expressed proteins, mutually confirmed each other in a way that seems to exclude the possibility the amino acid sequence variability of Fra a 1 has been largely underestimated. Seven positions of variability in the deduced amino acid sequence were suggested from cloning. All these seven positions were confirmed by mass spectrometry, and for three of the positions, both the variable amino acids were indeed detected in the expressed protein by mass spectrometry. Mass spectrometry detected only one more position of variability (both Val and Leu detected at position 24; Hjermø et al. (2006), which was “missed” by the cloning procedure (which detected only Leu24) with the presently used pair of primers.

Isoforms of proteins, appearing as distinct spots in 2DE, may be due to allelic variation causing amino acid sequence variability, or to post-translational modifications. The importance of protein post-translational modification (PTM) is becoming increasingly recognized with the possibility to detect and quantify PTMs by mass spectrometry (Jensen, 2004; Larsen and Roepstorff, 2000; Mann and Jensen, 2003). PTMs can affect protein activity, localization, half-life, interaction with other macromolecules, etc. such that differences between any two cellular states can be manifested by changes in PTMs rather than changes in protein expression. Since there appears to be so little variability in the Fra a 1 amino acid sequence, the isoforms detected in 2DE are likely to be due to post-translational modi-

fications. One possibility is that the different spots can represent differently phosphorylated forms, since there are several potential phosphorylation sites in the Fra a 1 sequence, as predicted by the NetPhos2.0 server (Blom et al., 1999). In order to further investigate if post-translational modifications can explain the appearance of the Fra a 1 allergen as four distinct spots in 2DE we will use the recombinantly expressed Fra a 1 to isolate Fra a 1-specific Ig G from polyclonal antiserum, and subsequently achieve affinity enrichment of nFra a 1 from strawberry. It is likely that it will be possible to find hypoallergenic strawberries by determining differences in Fra a 1, in terms of quantity and PTMs, from different varieties and growth conditions.

Acknowledgments

This study was supported by grant to A.M.-P. from Lars Hiertas Minne Foundation and to C.E. from the Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning Dnr 225-2004-1790 and Magnus Bergvalls Stiftelse.

References

- Aalberse, R.C., Akkerdaas, J., van Ree, R., 2001. Cross-reactivity of IgE antibodies to allergens. *Allergy* 56, 478–490.
- Bjerrum, O.J., Schafer Nielsen, C., 1986. Buffer systems and transfer parameters for semidry electroblotting with a horizontal apparatus. In: Dunn, M.J. (Ed.), *Electrophoresis-86; Fifth Meeting of the International Electrophoresis Society*. VCH, Weinheim, Germany, London, pp. 315–327.
- Blom, N., Gammeltoft, S., Brunak, S., 1999. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J. Mol. Biol.* 294, 1351–1362.
- Breiteneder, H., Ebner, C., 2000. Molecular and biochemical classification of plant-derived food allergens. *J. Allergy Clin. Immunol.* 106, 27–36.
- Ferreira, F., Hirtenlehner, K., Jilek, A., Godnik-Cvar, J., Breiteneder, H., Grimm, R., Hoffmann-Sommergruber, K., Scheiner, O., Kraft, D., Breitenbach, M., Rheinberger, H.J., Ebner, C., 1996. Dissection of immunoglobulin E and T lymphocyte reactivity of isoforms of the major birch pollen allergen Bet v 1: potential use of hypoallergenic isoforms for immunotherapy. *J. Exp. Med.* 183, 599–609.
- Gao, Z.S., van de Weg, W.E., Schaart, J.G., Schouten, H.J., Tran, D.H., Kodde, L.P., van der Meer, I.M., van der Geest, A.H., Kodde, J., Breiteneder, H., Hoffmann-Sommergruber, K., Bosch, D., Gilissen, L.J., 2005. Genomic cloning and linkage mapping of the Mal d 1 (PR-10) gene family in apple (*Malus domestica*). *Theor. Appl. Genet.* 111, 171–183.
- Gobom, J., Nordhoff, E., Mirgorodskaya, E., Ekman, R., Roepstorff, P., 1999. Sample purification and preparation technique based on nano-scale reversed-phase columns for the sensitive analysis of complex peptide mixtures by matrix-assisted laser desorption/ionization mass spectrometry. *J. Mass Spectrom.* 34, 105–116.
- Goodall, G.J., Filipowicz, W., 1991. Different effects of intron nucleotide composition and secondary structure on pre-mRNA splicing in monocot and dicot plants. *EMBO J.* 10, 2635–2644.
- Harndahl, U., Tufvesson, E., Sundby, C., 1998. The chloroplast small heat shock protein—purification and characterization of pea recombinant protein. *Protein Expr. Purif.* 14, 87–96.
- Hjernø, K., Alm, R., Canbäck, B., Matthiesen, R., Trajkovski, K., Björk, L., Roepstorff, P., Emanuelsson, C.S., 2006. Down-regulation of the strawberry Bet v 1-homologous allergen in concert with the flavonoid biosynthesis pathway in colourless strawberry mutant. *Proteomics* 6, 1574–1587.
- Hoffmann-Sommergruber, K., Vanek-Krebitz, M., Radauer, C., Wen, J., Ferreira, F., Scheiner, O., Breiteneder, H., 1997. Genomic characterization of members of the Bet v 1 family: genes coding for allergens and pathogenesis-related proteins share intron positions. *Gene* 197, 91–100.
- Horton, R.M., Ho, S.N., Pullen, J.K., Hunt, H.D., Cai, Z., Pease, L.R., 1993. Gene splicing by overlap extension. *Meth. Enzymol.* 217, 270–279.
- Hsieh, L.S., Moos Jr., M., Lin, Y., 1995. Characterization of apple 18 and 31 kd allergens by microsequencing and evaluation of their content during storage and ripening. *J. Allergy Clin. Immunol.* 96, 960–970.
- Jensen, O.N., 2004. Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. *Curr. Opin. Chem. Biol.* 8, 33–41.
- Karamloo, F., Scheurer, S., Wangorsch, A., May, S., Hausteine, D., Vieths, S., 2001. Pyr c 1, the major allergen from pear (*Pyrus communis*), is a new member of the Bet v 1 allergen family. *J. Chromatogr. B. Biomed. Sci. Appl.* 756, 281–293.
- Karlsson, A.L., Alm, R., Ekstrand, B., Fjellner-Modig, S., Schiott, A., Bengtsson, U., Björk, L., Hjernø, K., Roepstorff, P., Emanuelsson, C.S., 2004. Bet v 1 homologues in strawberry identified as IgE-binding proteins and presumptive allergens. *Allergy* 59, 1277–1284.
- Larsen, M.R., Roepstorff, P., 2000. Mass spectrometric identification of proteins and characterization of their post-translational modifications in proteome analysis. *Fresenius J. Anal. Chem.* 366, 677–690.
- Mann, M., Jensen, O.N., 2003. Proteomic analysis of post-translational modifications. *Nat. Biotechnol.* 21, 255–261.
- Marzban, G., Puehringer, H., Dey, R., Brynda, S., Ma, Y., Martinelli, A., Zaccarini, M., van der Weg, E., Housley, Z., Kolarich, D., Altmann, F., Laimer, M., 2005. Localization and distribution of the major allergens in apple fruit. *Plant Sci.* 169, 387–394.
- Mills, E., Madsen, C., Shewry, P., Wichers, H., 2003. Food allergens of plant origin—their molecular and evolutionary relationships. *Trends Food Sci. Technol.* 14, 145–156.
- Rappsilber, J., Ishihama, Y., Mann, M., 2003. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* 75, 663–670.
- Reuter, A., Fortunato, D., Garoffo, L.P., Napolitano, L., Scheurer, S., Giuffrida, M.G., Vieths, S., Conti, A., 2005. Novel isoforms of Pru av 1 with diverging immunoglobulin E binding properties identified by a synergistic combination of molecular biology and proteomics. *Proteomics* 5, 282–289.
- Son, D.Y., Scheurer, S., Hoffmann, A., Hausteine, D., Vieths, S., 1999. Pollen-related food allergy: cloning and immunological analysis of isoforms and mutants of Mal d 1, the major apple allergen, and Bet v 1, the major birch pollen allergen. *Eur. J. Nutr.* 38, 201–215.
- Swoboda, I., Jilek, A., Ferreira, F., Engel, E., Hoffmann-Sommergruber, K., Scheiner, O., Kraft, D., Breiteneder, H., Pittenauer, E., Schmid, E., et al., 1995a. Isoforms of Bet v 1, the major birch pollen allergen, analyzed by liquid chromatography, mass spectrometry, and cDNA cloning. *J. Biol. Chem.* 270, 2607–2613.
- Swoboda, I., Scheiner, O., Heberlebers, E., Vicente, O., 1995b. cDNA cloning and characterization of 3 genes in the Bet-V-1 gene family that encode pathogenesis-related proteins. *Plant Cell Environ.* 18, 865–874.
- Vieths, S., Scheurer, S., Ballmer-Weber, B., 2002. Current understanding of cross-reactivity of food allergens and pollen. *Ann. NY Acad. Sci.* 964, 47–68.
- Vieths, S., Schoning, B., Petersen, A., 1994. Characterization of the 18 kDa apple allergen by two-dimensional immunoblotting and microsequencing. *Int. Arch. Allergy Immunol.* 104, 399–404.